EXAM 01 PRACTICE QUESTIONS
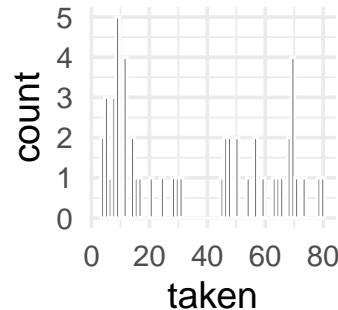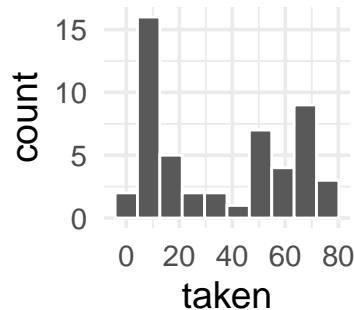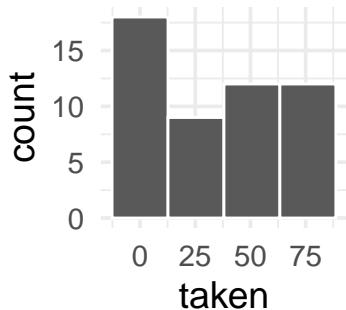
Stat 120 | Fall 2025
Prof Amanda Luby
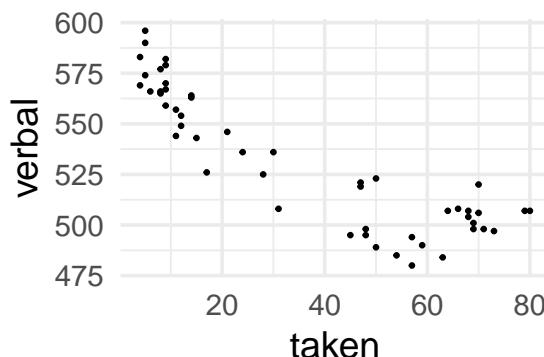
This dataset gives education-related data for the 50 states and the District of Columbia. The variables are:

- region: West, Northeast, Midwest, South
- pop: Population, in 1,000's
- verbal and math: average SAT verbal and math scores
- taken: percent of students taking the SAT
- noHS: percent of population with no high school diploma
- teachersPay: median teacher salary, in 1,000's

a) Label each variable as categorical or quantitative
b) How many columns and rows are there in this dataset?
c) Is this an experiment or observational study?
d) Below is the distribution of taken represented by 3 histograms. Choose the histogram that best represents the data and then describe the distribution.

e) Below is the scatterplot of verbal against taken. What do you think is closest to the actual correlation? Choose one: -1.1, -1, -.9, -.6, -.1, .1, .6, .9, 1, 1.1

f) What do the points in the *upper left* part of the graph represent? What about the *lower right*?
g) Below is the linear regression model output for this data. Write out the linear regression equation and interpret the slope and intercept in context.
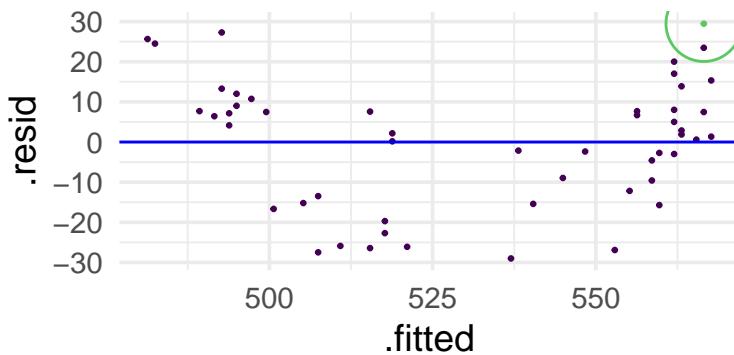
```
Call:
lm(formula = verbal ~ taken, data = sat)

Coefficients:
(Intercept)        taken
    572.21        -1.14
```
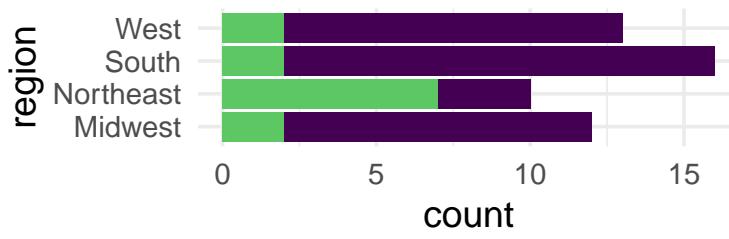
h) Below is the residual plot for this model. Do you have any concerns about the validity of the model?

i) One point on the residual plot is colored in green and circled. Give the $\hat{y}$, ϵ, y, and x value for this data point.

j) The mean of teachersPay is 35.89, the median is 35, and the standard deviation is 6.226. The distribution is approximately symmetric and bell-shaped. Approximately how many states have teachersPay between 29.66 and 42.12?

k) The graph below displays region filled by a new variable, pay40: whether teachersPay is greater than 40.

k) Find the conditional distribution of pay40 for the Northeast region

l) Estimate the proportion of Midwest states with a median teacher pay less than 40.
m) Find the marginal distribution of `pay40`
n) Does there appear to be a relationship between `pay40` and `region`?

There will also be a few *conceptual* true/false or multiple choice questions that are *not* connected to the data analysis.